

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-249676

(43) 公開日 平成11年(1999) 9月17日

(51) Int.Cl.⁶

G 1 0 L 3/00
5/04

識別記号

F I

G 1 0 L 3/00
5/04

H
F

審査請求 未請求 請求項の数 6 O L (全 12 頁)

(21) 出願番号 特願平10-47471

(22) 出願日 平成10年(1998) 2月27日

(71) 出願人 000108085

セコム株式会社

東京都新宿区西新宿 1 丁目26番 2 号

(72) 発明者 濱上 知樹

東京都三鷹市下連雀 8 丁目10番16号 セコム株式会社内

(72) 発明者 古村 光夫

東京都三鷹市下連雀 8 丁目10番16号 セコム株式会社内

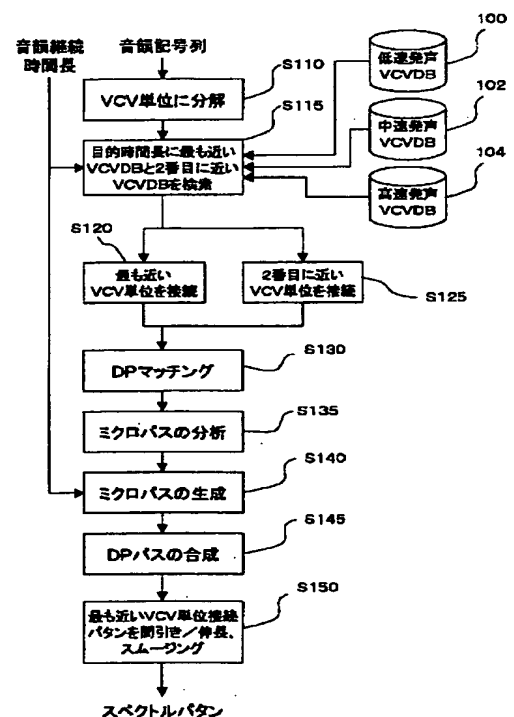
(74) 代理人 弁理士 吉田 研二 (外 2 名)

(54) 【発明の名称】 音声合成装置

(57) 【要約】

【課題】 音声合成装置において、スペクトルパタンのデータベースサイズを抑制しつつ、任意の発話速度に対して自然な音韻を実現する。

【解決手段】 異なる代表時間長に対応したスペクトルパタンを格納する複数のデータベース100~104から、目的継続時間長に近い時間長のものを2つ選択する。それらから取り出した2つのスペクトルパタン間の非線形の対応関係をDPマッチングにより求める(S130)。この対応関係を分析して線形成分であるマクロバス、残りの非線形成分であるマイクロバスが得られる(S135)。目的時間長に対するマクロバスは時間長の比を傾きとする直線として求められる。目的時間長に対するマイクロバスは、対応関係から得られる各点での値を補間して求められる(S140)。これらを合成したDPバスに基づいて、基準スペクトルパタンの非線形伸縮が行われ、目的時間長のスペクトルパタンが得られる。



【特許請求の範囲】

【請求項 1】 テキストから音韻記号列を生成し、前記音韻記号列を区分した音韻区画毎にスペクトルボタンを生成し、音声規則合成する音声合成装置において、同一の前記音韻区画に対する異なる代表継続時間長に対応した複数の代表時間長スペクトルボタンそれぞれの時間変化同士の対応関係に基づき、いずれかの前記代表時間長スペクトルボタンの前記時間変化を目的継続時間長に応じて非線形伸縮変換して、前記目的継続時間長に対応した目的時間長スペクトルボタンを生成するスペクトルボタン生成手段を有し、
当該目的時間長スペクトルボタンを用いて音声合成することを特徴とする音声合成装置。

【請求項 2】 請求項 1 記載の音声合成装置において、前記スペクトルボタン生成手段は、前記音韻区画の種類毎に前記複数の代表時間長スペクトルボタンを格納したスペクトルボタン記憶手段と、前記目的継続時間長に応じて前記複数の代表時間長スペクトルボタンのいずれかを選択する基準スペクトルボタン選択手段と、前記スペクトルボタン記憶手段に格納された前記代表時間長スペクトルボタンに基づいて、前記各音韻区画毎に前記複数の代表時間長スペクトルボタンそれぞれの時間変化同士の前記対応関係を生成する対応関係生成手段と、
前記対応関係に基づく前記音韻区画内での非線形伸縮変換を、前記目的継続時間長に応じて前記基準スペクトルボタンに施す非線形伸縮手段と、
を有することを特徴とする音声合成装置。

【請求項 3】 請求項 1 記載の音声合成装置において、前記スペクトルボタン生成手段は、前記音韻区画の種類毎に前記複数の代表時間長スペクトルボタンを格納したスペクトルボタン記憶手段と、前記目的継続時間長に応じて前記複数の代表時間長スペクトルボタンのいずれかを選択する基準スペクトルボタン選択手段と、
前記音韻区画の種類毎に前記複数の代表時間長スペクトルボタンそれぞれの時間変化同士の前記対応関係を格納した対応関係記憶手段と、
前記対応関係に基づく前記音韻区画内での非線形伸縮変換を、前記目的継続時間長に応じて前記基準スペクトルボタンに施す非線形伸縮手段と、
を有することを特徴とする音声合成装置。

【請求項 4】 請求項 1 記載の音声合成装置において、前記スペクトルボタン生成手段は、前記音韻区画の種類毎の前記複数の代表時間長スペクトルボタンのうち基準となる基準スペクトルボタンを格納したスペクトルボタン記憶手段と、
前記音韻区画の種類毎に前記基準スペクトルボタンと他の前記代表時間長スペクトルボタンと間での時間変化の

前記対応関係を格納した対応関係記憶手段と、
前記対応関係に基づく前記音韻区画内での非線形伸縮変換を、前記目的継続時間長に応じて前記基準スペクトルボタンに施す非線形伸縮手段と、
を有することを特徴とする音声合成装置。

【請求項 5】 請求項 1 から請求項 4 のいずれかに記載の音声合成装置において、
前記複数の代表時間長スペクトルボタンそれぞれの時間変化同士の前記対応関係は、当該代表時間長スペクトルボタンの継続時間長に比例した線形伸縮成分と、当該線形伸縮成分を除いた非線形伸縮成分とに分解され、
前記非線形伸縮変換は、
前記線形伸縮成分を前記継続時間長に応じて補間して補間線形伸縮成分を求め、
前記非線形伸縮成分を前記継続時間長に応じて補間して補間非線形伸縮成分を求め、
前記補間線形伸縮成分と前記補間非線形伸縮成分とを合成すること、
を特徴とする音声合成装置。

【請求項 6】 請求項 1 から請求項 5 のいずれかに記載の音声合成装置において、
前記複数の代表時間長スペクトルボタンそれぞれの時間変化同士の前記対応関係は、DP マッチング（動的計画法）により求められることを特徴とする音声合成装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は音声規則合成する音声合成装置、特に自然で滑らかな音韻変化を生成し合成音声の品質を向上する音声合成装置に関する。

【0002】

【従来の技術】 音声合成の方式は、再生方式、分析合成方式、及び規則合成方式の 3 つに大きく分類できる。再生方式は、予め記録した単語単位の音声波形を接続して再生する方式である。分析合成方式は、例えば単語や文といった単位のある程度まとまった音声データを予め分析し、音声の冗長性を取り除き、圧縮された形のパラメータで蓄えておき、必要に応じて合成装置にて音声を合成する方式である。規則合成方式は、単語構成、文型などによって多様に変化する文の韻律的特徴を規則により表現する高度な合成方式であり、細かな音声単位を基に、これを組合せ、発話速度（又は継続時間長）に応じて個々の単位パラメータの長さを調整し、変形、アクセントやイントネーション、パワーの変化を規則によって導出する。この方式によれば任意の文字から高品質の合成音声を生成することができる。

【0003】 音声は、音色、基本周波数の高さ、及び強度の 3 要素に分解して捉えることができる。つまり、これら 3 要素を決定することにより音声を合成することができる。スペクトルボタンは、これらのうち音色を表す情報である。スペクトルボタンは、日本語テキスト（漢

字仮名交じり文)を構成する各音韻の周波数スペクトルの包絡パタンを、日本語テキストから生成される音韻記号列に従った順序で結合して生成される。このスペクトルパタンは、合成音声の音韻性そのものや滑らかさに影響を及ぼすため、その生成は規則合成方式における重要な技術の1つである。

【0004】図8は、従来のスペクトルパタン生成方式における処理の流れを示す模式図である。従来より行われているスペクトルパタンの生成方法は、日本語テキストから生成された音韻記号列をVCV音節といった所定の音節単位に分解する(S10)。次に当該音節単位毎のスペクトルパタンをデータベース15から検索し(S20)、これらを滑らかに結合した後(S25)、各音節単位のスペクトルパタンを当該音節単位の音韻継続時間長に応じて線形に伸縮する(S30)というものであった。ここで用いられる音節単位は、母音-子音-母音の連鎖パタンであるVCV音節や、子音-母音の連鎖パタンであるCV音節である。

【0005】また、他の従来方法として、例えば、「音韻環境に基づくクラスタリングによる規則合成法」(中嶋 他、電子情報通信学会論文誌D-II Vol. J72-D-II No. 8pp. 1174、1989年8月)に示されるような合成単位を一定の音節単位ではなく可変としたもの(可変単位合成方式と称する。)もあった。この方式は、結合要素となる音韻単位の種類を例えばVCV音節といった一定の連鎖パタンに限ることなく、発話速度や音韻環境条件に応じて拡張したデータベースを構築することにより、伸縮操作なしに自然性の高いスペクトルパタンを得ようとするものである。また、他の可変単位合成方式としては、例えば、「動的特徴を用いたHMMに基づく音声合成」(益子 他、電子情報通信学会論文誌D-II Vol. J79-D-II No. 12 pp. 2184、1996年12月)に示されるような、スペクトルパタンの動的特徴を抽出・蓄積し、スペクトルパタンの生成においてその動的特徴を反映させることにより自然性の向上を図るものも提案されている。

【0006】

【発明が解決しようとする課題】音節単位のデータベースを用いた従来の合成方法においては、与えられた音節単位の継続時間長に対しては自然な音韻を実現するスペクトルパタンが得られる。しかし、音節単位の継続時間長は発話速度によって変化し、その他、音韻環境や隣接音韻の影響による音韻継続時間長の変動などによっても変化する。従来の音節単位のデータベースを用いたスペクトルパタンの合成方法は、このような変化に対応した音節単位の任意の継続時間長に対するスペクトルパタンを、音節単位毎の所与の継続時間長のスペクトルパタンを当該単位内で単純に線形伸縮して生成するため、上記要因等による音節単位の継続時間長の変化に伴うスペクトルパタン変化の推移が不自然になるおそれがあるといった問題を有していた。つまり、例えば、文章全体又は

部分で発話速度を変化させる場合に、自然な音韻性が得られないという問題や、音韻環境や隣接音韻の影響を補償した音韻継続時間長を生成しても、その補償効果を十分に音韻の自然さに反映することが容易でないという問題があった。

【0007】また、音韻単位の種類を拡大する可変単位合成方式では、データベースの規模が大きくなるといった問題や、単位同士の接続組合せパタンが増えることにより、各スペクトルパタンの端部の調整が難しくなり接続部で歪みが発生しやすくなるという問題や、特徴を抽出するための学習データの増加に伴い学習時間がかかるといった問題があった。

【0008】また、伸縮規則を適用する区間(セグメンテーション)を決定するためのラベリングを人手により行うことは、その者の主観に依存する部分も大きく、区間の定義が不正確、不安定になる。そのため、例えば異なる者のラベリングに基づいて作成された規則同士は異なるものとなり、普遍性がないという問題がある。例えば、音節単位といった比較的大きな単位では人に依存する誤差は小さいと考えられるが、より微細な区間についてのラベリングでは上記問題が顕著になるおそれがある。

【0009】本発明は上記問題を解決することを目的としたものであり、音韻単位のデータベースを用いた合成方式において、データベースの規模を抑制し、かつ学習を行うことなく、音節単位の接続歪みが生じにくく、自然な合成音声を得ることができる音声合成装置を提供することを目的とする。

【0010】

【課題を解決するための手段】本発明に係る音声合成装置は、同一の音韻区画に対する異なる代表継続時間長に対応した複数の代表時間長スペクトルパタンそれぞれの時間変化同士の対応関係に基づき、いずれかの前記代表時間長スペクトルパタンの前記時間変化を目的継続時間長に応じて非線形伸縮変換して、前記目的継続時間長に対応した目的時間長スペクトルパタンを生成するスペクトルパタン生成手段を有し、当該目的時間長スペクトルパタンを用いて音声合成するものである。

【0011】本発明によれば、音韻記号列を区分した音韻区画毎にスペクトルパタンが生成され、接続され、そのスペクトルパタンを用いて音声規則合成される。音韻区画は、例えばVCV音節といった音節単位である。同一の音韻記号列を含んだ音韻区画を異なる継続時間長、すなわち異なる発話速度、音韻継続時間長で発声した場合、一般にそれら異なる継続時間長に対応するスペクトルパタンは互いに時間軸を継続時間長に応じて単純に比例伸縮して得られる相似形とはならない。つまり、継続時間長が異なった場合、一般にスペクトルパタンの時間的变化(ラベル付けされた各点)の順序は維持されるが、スペクトルパタンの各部分は時間軸方向に必ずし

も一様に伸縮しない。本発明はこの点に鑑み、複数の代表的な継続時間長（代表継続時間長）それぞれについての同一音韻区画のスペクトルボタン（代表時間長スペクトルボタン）間での時間変化同士の対応関係に基づいて、所望の目的継続時間長に対応した目的時間長スペクトルボタンを生成する。

【0012】複数のスペクトルボタンの時間変化同士の対応関係は、代表時間長スペクトルボタン間での時間軸の変換関係を表し、例えば、第一の代表時間長スペクトルボタンのその先頭からある時刻 t_1 におけるスペクトルが、第二の代表時間長スペクトルボタンのどの時刻 t_2 に現れるかというものである。例えば2つの代表時間長スペクトルボタン間の対応関係に基づく場合、例えば目的継続時間長が第一の代表継続時間長と第二の代表継続時間長との間の値であるときは、第一の代表時間長スペクトルボタンのある時刻 t_1 に対応するスペクトルが目的時間長スペクトルボタン上に現れる時刻 t は、例えば t_1 と t_2 との間の値に定めることができ、時刻 t は目的継続時間長が第一の代表継続時間長に近いほど、時刻 t_1 に近くなり、目的継続時間長が第二の代表継続時間長に近いほど、時刻 t_2 に近くなる。時刻 t を定めるために用いられる時間変化の対応関係は、一般には2つ以上の異なる代表継続時間長に対応した代表時間長スペクトルボタン相互間のものであり、必ずしも2つの代表時間長スペクトルボタン間のものに限られない。2つの代表時間長スペクトルボタンの対応時刻 t_1 、 t_2 から目的時間長スペクトルボタンの対応時刻 t を定める方法には、例えば目的継続時間長に応じた t_1 、 t_2 の線形補間がある。また、3つ以上の代表時間長スペクトルボタンの対応時刻 t_1 、 t_2 、 \dots 、 t_i から目的時間長スペクトルボタンの対応時刻 t を定める場合には、例えば高次関数を用いた補間を用いることができる。

【0013】代表時間長スペクトルボタン間の時間変化の対応関係から、代表時間長スペクトルボタンの時間軸から目的時間長スペクトルボタンの時間軸への変換関係が定められる。目的時間長スペクトルボタンは、この変換関係に基づいて、ある代表時間長スペクトルボタンを時間軸方向に関して伸縮変換することにより得られる。上述したように一般にある継続時間長の代表時間長スペクトルボタンに対する他の継続時間長の代表時間長スペクトルボタンの伸縮が不均一であることから、目的時間長スペクトルボタンの時間軸の伸縮も不均一である。つまり、例えば、目的継続時間長が第一の代表継続時間長の α 倍になった場合、伸縮が一樣であれば第一の代表時間長スペクトルボタンの時刻 t_1 と目的時間長スペクトルボタンの時刻 t との間には $t = \alpha \cdot t_1$ という線形関係が成り立つのであるが、一般には上記不均一性によりこのような線形関係は成り立たず、代表時間長スペクトルボタンと目的時間長スペクトルボタンとは非線形の伸縮変換により関係付けられる。本発明はこの一般的な場

合である非線形伸縮変換に対応することにより、滑らかで自然な合成音声を得るという目的を達成する。

【0014】本発明の好適な態様は、前記スペクトルボタン生成手段が、前記音韻区画の種類毎に前記複数の代表時間長スペクトルボタンを格納したスペクトルボタン記憶手段と、前記目的継続時間長に応じて前記複数の代表時間長スペクトルボタンのいずれかを選択する基準スペクトルボタン選択手段と、前記スペクトルボタン記憶手段に格納された前記代表時間長スペクトルボタンに基づいて、前記各音韻区画毎に前記複数の代表時間長スペクトルボタンそれぞれの時間変化同士の前記対応関係を生成する対応関係生成手段と、前記対応関係に基づく前記音韻区画内での非線形伸縮変換を、前記目的継続時間長に応じて前記基準スペクトルボタンに施す非線形伸縮手段とを有するものである。また本発明の他の好適な態様は、上記態様における対応関係生成手段に代えて、前記音韻区画の種類毎に前記複数の代表時間長スペクトルボタンそれぞれの時間変化同士の前記対応関係を格納した対応関係記憶手段を有するものである。これらの態様において、基準スペクトルボタンは、例えば、その継続時間長が目的継続時間長に近いものを採用するのが好適である。

【0015】本発明の別の好適な態様は、前記スペクトルボタン生成手段が、前記音韻区画の種類毎の前記複数の代表時間長スペクトルボタンのうち基準となる基準スペクトルボタンを格納したスペクトルボタン記憶手段と、前記音韻区画の種類毎に前記基準スペクトルボタンと他の前記代表時間長スペクトルボタンと間での時間変化の前記対応関係を格納した対応関係記憶手段と、前記対応関係に基づく前記音韻区画内での非線形伸縮変換を、前記目的継続時間長に応じて前記基準スペクトルボタンに施す非線形伸縮手段とを有するものである。本態様において、特に基準スペクトルボタンには、音韻情報の欠落が少ない発話速度が遅いものを採用するのが好適である。

【0016】他の本発明に係る音声合成装置においては、前記複数の代表時間長スペクトルボタンそれぞれの時間変化同士の前記対応関係は、当該代表時間長スペクトルボタンの前記継続時間長に比例した線形伸縮成分と、当該線形伸縮成分を除いた非線形伸縮成分とに分解され、前記非線形伸縮変換は、前記線形伸縮成分を前記継続時間長に応じて補間して補間線形伸縮成分を求め、前記非線形伸縮成分を前記継続時間長に応じて補間して補間非線形伸縮成分を求め、前記補間線形伸縮成分と前記補間非線形伸縮成分とを合成する。

【0017】本発明に係る音声合成装置は、前記複数の代表時間長スペクトルボタンそれぞれの時間変化同士の前記対応関係が、DPマッチング（動的計画法）により求められるものである。

【0018】本発明によれば、DPマッチングを用いて

対応関係を定めることにより、前記代表時間長スペクトルパタンのある時刻に対応する他のスペクトルパタン上での時刻が客観的・定量的に決定される。

【0019】

【発明の実施の形態】〔実施の形態1〕以下、本発明の実施の形態を図面を参照して説明する。図1は本実施形態に係る日本語音声合成装置のブロック構成図である。本音声合成装置は従来同様に、電子化された日本語テキストの文字データから合成音声規則を合成して出力するものである。

【0020】テキスト解析処理部52は、入力された日本語テキストの言語解析結果を出力する。この解析結果には、音声合成処理に必要なアクセントの情報、フレーズの境界、ポーズ、母音の無声化といった発音情報を加えた音韻記号列や、フレーズやアクセント句の分類結果が含まれている。

【0021】この言語解析結果は音韻継続時間長生成部54に入力される。音韻継続時間長生成部54は、目的の発話速度を実現する音韻の時間長情報を音韻記号列に基づいて音節や音素に付加し、これをスペクトルパタン生成部56、音源振幅パタン生成部58及びピッチパタン生成部60に出力する。

【0022】本実施形態の特徴的構成要素であるスペクトルパタン生成部56は、音韻結合規則により、母音・子音といった音韻の種類からデータベースを検索し、各音韻のスペクトルパタンを結合し、一連の新たなスペクトルパタンを生成して出力するものであり、これについては後に詳述する。

【0023】音源振幅パタン生成部58は、音声のパワー包絡をパワーに関する規則により決定する。ピッチパタン生成部60は、韻律制御についての規則からピッチパタンを生成するものである。

【0024】音源生成部62は、音源振幅パタン生成部58からのパワーパタンと、ピッチパタン生成部60からのピッチパタンとを入力とし、音源情報を生成する。音声合成部64は、スペクトルパタン生成部56から入力されるスペクトルパタンによって、音源生成部62からの音源情報を変調して音色を付加し合成音声を生成する。

【0025】図2は、本実施形態に係る音声合成装置のスペクトルパタン生成部56の処理を説明する模式図である。

【0026】スペクトルパタン生成部56は、VCV単位の各種類についてそれぞれ3種類の代表的な発話速度のスペクトルパタン（代表時間長スペクトルパタン）をデータベースに格納・保持している。例えば、代表的な発話速度は、標準的な速度である7～8mora/秒、及びその前後の速度を用いることが好ましい。例えば、低速、中速、高速の各発話速度として6mora/秒、8mora/秒、11mora/秒を用いることができる。図では各速度

に対応する代表時間長スペクトルパタンはそれぞれ低速発声VCVデータベース100、中速発声VCVデータベース102、高速発声VCVデータベース104に格納されている。ここで、VCV音節は、例えば両端の母音部重心位置を端点とするように定義される。

【0027】スペクトルパタン生成部56には、音韻継続時間長生成部54から音韻継続時間長を付与された音韻記号列が入力される。音韻記号列はVCV単位に分割される（S110）。そしてVCV音節毎にその音韻継続時間長に最も近い発話速度に対応するデータベースと2番目に近い発話速度に対応するデータベースとを上記データベース100～104からそれぞれ選択し、各データベースから当該VCV音節に対応する代表時間長スペクトルパタンを検索し取り出す（S115）。次に、各VCV音節に対して選択された最も発話速度に近い代表時間長スペクトルパタン（以下、基準スペクトルパタンと呼ぶ。）同士の接続（S120）と、各VCV音節に対して選択された2番目に発話速度に近い代表時間長スペクトルパタン（以下、参照スペクトルパタンと呼ぶ。）同士の接続（S125）とを行い、最も近い発話速度と2番目に近い発話速度に対応したVCV単位接続スペクトルパタンをそれぞれ生成する。ちなみに、各VCV単位接続スペクトルパタン中では、異なる発話速度のVCV単位が接続されうる。つまり、例えば、最も近い発話速度のVCV単位に中速のVCV単位が接続されるということが起こりうる。

【0028】本装置のスペクトルパタン生成部56は、VCV音節をさらに細かな基本単位であるフレームに区分し、基準スペクトルパタンと参照スペクトルパタンとの各フレームの対応関係をDPマッチングにより求める（S130）。そして、音韻継続時間長生成部54から与えられる目的とする継続時間長のスペクトルパタンが、このDPマッチングにより得られた対応関係に基づいて生成される（S135～S150）。以下、このDPマッチングを用いた本装置のスペクトルパタン生成方法を詳しく述べる。

【0029】図3は、同一のVCV音節に対応した3種類の発話速度の代表時間長スペクトルパタンをそれぞれ示す模式図である。スペクトルパタン160は、低速の発話速度の代表時間長スペクトルパタンであり、その全体はI個のフレームに分割される。この低速に対応した代表時間長スペクトルパタンは、LPC（Linear Predictive Coding）フレーム分析によって、スペクトルパタンの系列 $Ss[i]$ として低速発声VCVデータベース100に登録されている。ここでiは先頭から数えたフレーム番号であり、 $1 \leq i \leq I$ である。同様に、スペクトルパタン162、164は、それぞれ中速、高速の発話速度の代表時間長スペクトルパタンであり、その全体はそれぞれJ、K個のフレームに分割され、スペクトルパタ

ンの系列 $S_m[j]$ 、 $S_f[k]$ として中速発声 VCV データベース 102、高速発声 VCV データベース 104 に登録されている。ここで j 、 k は先頭から数えたフレーム番号であり、 $1 \leq j \leq J$ 、 $1 \leq k \leq K$ である。なお、ここでは、各フレームの時間幅が均等である場合を説明することとし、その時間幅は例えば 10ms といった値である。

【0030】DP マッチングを行うことにより、2つの音声 x 、 y のスペクトルボタン $S_x[n]$ 、 $S_y[m]$ ($1 \leq n \leq X$ 、 $1 \leq m \leq Y$) の間の対応関係が DP バスにより表される。DP バス P_{xy} は、音声 x の時刻 ζ のスペクトルが音声 y のどの時刻 ξ のスペクトルと一致するかを表すものである。図 4 は、横軸に音声 x の時間軸、縦軸に音声 y の時間軸をとり、DP バスを示した模式図である。図において、DP バスは、 x 、 y の継続時間長の比 Y/X を傾きとする直線 170 におおよそ沿った折れ線 172 で表されている。ここでは、DP バスを次式で表すこ

$$P'_{xy}[p] = (P_{xy1}[p], (Y/X) P_{xy1}[p]) \quad \dots\dots\dots (1)$$

である。また、 p 軸上における DP バスとマクロバスと ※にて表される。
の差をマイクロバス (dP_{xy}) と呼ぶこととする。ミクロ 20
バスは次式で表され、また、図 4 において折れ線 174 ※

$$\begin{aligned} dP_{xy}[p] &= (dP_{xy1}[p], dP_{xy2}[p]) \\ &= P_{xy}[p] - P'_{xy}[p] \\ &= (P_{xy1}[p], P_{xy2}[p] - (Y/X) P_{xy1}[p]) \quad \dots\dots\dots (2) \end{aligned}$$

以上の手法により、DP バスは線形伸縮よりなるマクロバスと、非線形伸縮よりなるマイクロバスに分解される。★

$$P_{xy}[p] = P'_{xy}[p] + dP_{xy}[p] \quad \dots\dots\dots (3)$$

と表される。

【0036】このようにして同一の VCV 音節に対する異なる代表継続時間長に対応した代表時間長スペクトルボタン間の対応関係がマクロバス、マイクロバスとして与えられる。これらを用い、いずれかの代表時間長スペクトルボタンを基準スペクトルボタンとして、伸縮することにより当該 VCV 音節に対する任意の継続時間長に対応したスペクトルボタン (目的時間長スペクトルボタン) を生成することができる。例えば、基準スペクトルボタンと目的時間長スペクトルボタンとの間の DP バスのマクロバス (P'_{xt}) は、それぞれの継続時間長の比 ☆

$$\begin{aligned} P'_{xt}[p] &= (T/Y) P'_{xy}[p] \\ &= (P_{xy1}[p], (T/X) P_{xy1}[p]) \quad \dots\dots\dots (4) \end{aligned}$$

$$\begin{aligned} dP_{xt}[p] &= \{(X-T)/(X-Y)\} dP_{xy}[p] \\ &= (dP_{xy1}[p], dP_{xy2}[p]) \\ &= (P_{xy1}[p], \{(X-T)/(X-Y)\} \{P_{xy2}[p] - (Y/X) P_{xy1}[p]\}) \quad \dots\dots\dots (5) \end{aligned}$$

本装置は、データベース 100～104 に格納されたスペクトルボタン $S_s[i]$ 、 $S_m[j]$ 、 $S_f[k]$ と以上のスペクトルボタンの伸縮方法とにより、目的時間長スペクトルボタン $S_t[h]$ ($1 \leq h \leq T$) を以下の手順で生成する。なお、ここでは $1 \leq T \leq K$ であるものとする。

【0038】スペクトルボタン生成部 56 は、処理 S1 50

*ととする。なお、音声 x 、 y の端部では、DP バスと直線 170 とは交わる。

$$【0031】P_{xy} = (\zeta, \xi)$$

ここで p を DP バス上のノード番号とし、これを媒介変数として ζ 、 ξ を次式で表すことにする。

$$【0032】\zeta = P_{xy1}[p]$$

$$\xi = P_{xy2}[p]$$

すると、DP バスは次式で表される。

$$【0033】P_{xy}[p] = (P_{xy1}[p], P_{xy2}[p])$$

一方、直線 170 は、音声 x と y の線形に伸縮させた、つまり全体を均一に伸縮させた場合のバスに相当し、ここではこれをマクロバス (P'_{xy}) と呼び、次式で表す。

$$【0034】P'_{xy}[p] = (P'_{xy1}[p], P'_{xy2}[p])$$

上述したようにマクロバスは、傾き Y/X の直線であるので、

$$\dots\dots\dots (1)$$

※にて表される。

$$【0035】$$

20

★すなわち、

$$\dots\dots\dots (3)$$

☆により定まる直線であるので、当該継続時間長比から直接的に求めることができる。一方、マイクロバス (dP_{xt}) は、まず自然音声代表時間長スペクトルボタンとして用いて求めたマイクロバス dP_{xy} を、継続時間長差の比によってスケールリングすることにより定めることができる。目的時間長スペクトルボタンを $S_t[h]$ ($1 \leq h \leq T$)、基準スペクトルボタンを例えば $S_x[n]$ とすると、これら目的時間長スペクトルボタンと基準スペクトルボタンとに関するマクロバス P'_{xt} 、マイクロバス dP_{xt} は、それぞれ次式で表される。

$$【0037】$$

$$\begin{aligned} P'_{xt}[p] &= (T/Y) P'_{xy}[p] \\ &= (P_{xy1}[p], (T/X) P_{xy1}[p]) \quad \dots\dots\dots (4) \end{aligned}$$

$$\begin{aligned} dP_{xt}[p] &= \{(X-T)/(X-Y)\} dP_{xy}[p] \\ &= (dP_{xy1}[p], dP_{xy2}[p]) \\ &= (P_{xy1}[p], \{(X-T)/(X-Y)\} \{P_{xy2}[p] - (Y/X) P_{xy1}[p]\}) \quad \dots\dots\dots (5) \end{aligned}$$

20 により継続時間長 D_x ($x = s, m, f$) である基準スペクトルボタン $S_x[n]$ ($1 \leq n \leq N$) を選択し、処理 S125 により継続時間長 D_y ($y = s, m, f$ 、但し $y \neq x$) である参照スペクトルボタン $S_y[m]$ ($1 \leq m \leq M$) を選択する。

【0039】次に、DP マッチングにより $S_x[n]$ に対す

る $Sy[m]$ の DP バス Pxy を作成する (S 1 3 0)。そして上述手順の (2) 式により、DP バス Pxy からマイクロバスを求め (S 1 3 5)、目的時間長スペクトルボタン $St[h]$ に関するマイクロバス $dPxt$ を (5) 式により生成する (S 1 4 0)。一方、マクロバス $P'xt$ は、上述したように継続時間長比から直接的に求めることができ、これと処理 S 1 4 0 で求めたマイクロバスとを合成して、DP バス Pxt が生成される (S 1 4 5)。ちなみに、このようにマイクロバスとマクロバスとを分解することにより、処理上、各点に対してマイクロバスのデータのみを例 * 10

$$\{(X-T)/(X-Y)\} Pxy2[p] + \{(T-Y)/(X-Y)\} Pxy1[p]\}$$

..... (6)

を用いて目的時間長スペクトルボタンに対する DP バスを求めることができる。ちなみに (6) 式は、代表時間長スペクトルボタン $Sx[n]$ 、 $Sy[m]$ の互いに対応する点 $Pxy1[p]$ 、 $Pxy2[p]$ に対応する目的時間長スペクトルボタンの $St[h]$ の点は、 $Pxy1[p]$ 、 $Pxy2[p]$ を各継続時間長に応じて内分した点であることを示している。

【0 0 4 1】さて、ここで、求められた DP バス上の各点は、それを表す座標を四捨五入することにより、フレームの格子に吸着させられる。つまり DP バスの各点はフレームを単位として表される。

【0 0 4 2】この DP バスのフレーム格子への吸着により、基準スペクトルボタンのあるフレームに対応する目的時間長スペクトルボタンのフレームがない (フレームの欠落) といったことや、基準スペクトルボタンのあるフレームに目的時間長スペクトルボタンの複数のフレームに対応する (フレームの伸長) といったことが起こりうる。よって DP マッチングによって得られたバスを基に、V C V 音節スペクトルボタンの伸縮を行うためには、フレームが欠落する箇所やフレームが伸長している箇所において不連続が生じないように補間操作を行う必要がある。本装置は以下の手順で、スペクトルの連続性を保ちながら、欠落したフレームに対応するスペクトルボタン部分を目的時間長スペクトルボタンから間引き、また伸長したフレームに対応するスペクトルボタン部分を接続し (S 1 5 0)、目的時間長スペクトルボタンを合成する。

【0 0 4 3】まず、合成された DP バス Pxt に基づいて、基準スペクトルボタン $Sx[n]$ ($1 \leq n \leq N$) の各フレームに、目的時間長スペクトルボタンのフレームとの対応関係に応じてラベルを付与する。目的時間長スペクトルボタンに対応するフレームがないとき、すなわちフレームの欠落が生じている箇所においては、消去ラベル “0” が付与される。また、基準スペクトルボタンのフレームに対して目的時間長スペクトルボタンのフレームが一つのみ対応付けられるときは、保存ラベル “1” を、一方、 k 個のフレームが対応付けられるとき、すなわち伸長が生じる箇所においては、伸長するフレーム数で表される伸長ラベル “ k ” が付与される。

* えばメモリ上に保持すればよい。これにより計算上、取り扱うパラメータを削減することができ処理上有利である。またマイクロバスのデータがとりうる範囲は、分解しない場合の DP バスのデータがとりうる範囲より狭くすることができるので、例えば保持すべきデータ量を削減したり、同一のビット数でより精度の高い数値を保持することができるといったメリットも得られる。

【0 0 4 0】なお、分解しない場合は、(4) (5) 式から得られる次式、 $Pxt[p] = (Pxy1[p],$

【0 0 4 4】保存ラベル “1” が付与されたフレームに対しては補間操作は行われぬ。伸長ラベル “ k ” が付与された基準スペクトルボタンのフレームに対する目的時間長スペクトルボタンの k 個のフレームに対しては、基準スペクトルボタンにおけるラベル “ k ” のフレームとこれに先行するフレームとを線形補間したフレーム、及び基準スペクトルボタンにおけるラベル “ k ” のフレームとこれに後続するフレームとを線形補間したフレームが生成され割り当てられる。

【0 0 4 5】消去ラベル “0” を付与されたフレームが拗音、鼻音を除く子音区間の場合、もともとスペクトルの連続性が低いので補間を行わずフレームを除去する。一方、消去ラベル “0” を付与されたフレームが母音、撥音、拗音、鼻音の場合、当該フレーム (ラベル “0” が連続する場合は、当該連続フレーム) を除去した後、当該フレームを含む音素 (V 又は C) 内のフレーム間で次に示す平滑化操作を行い、フレーム消去に伴うスペクトルの不連続を生じさせないようにする。

【0 0 4 6】ここで、スペクトルボタンは L S P (Line Spectrum Pair) パラメータを用いて表され、基準スペクトルボタンの n フレーム目の r 次の L S P パラメータを $Sx[n][r]$ 、平滑化後のパラメータを $S'x[n][r]$ と表記する。消去された結果、互いに接続する必要のある境界フレームを c 、 $c+d$ 、消去フレームを含む音素区間の開始フレーム、終了フレームをそれぞれ b 、 e とすると、以下に示す計算により平滑化後のパラメータを求めることができる。まず接続される c フレームと ($c+d$) フレームではパラメータは等しくなるという境界条件から、両境界フレームの平滑化後のパラメータの値に、平滑化前の c フレームと ($c+d$) フレームとのパラメータの平均値を与える。つまり、

$$S'x[c][r] \equiv S'x[c+d][r] \equiv (Sx[c][r] + Sx[c+d][r]) / 2$$

とする。そしてこの境界条件の下、 $b \sim c$ フレームと ($c+d$) $\sim e$ フレームとの各区間において、それぞれ重み関数 $w1[n]$ 、 $w2[m]$ を用いて平滑化を行う。具体的には $b \leq n < c$ なる n フレームにおいては

$$S'x[n][r] \equiv Sx[n][r] - (Sx[n+1][r] - S'x[n+1][r])$$

1][r]) w1[n]

また $c + d < m \leq e$ なる m フレームにおいては
 $S'x[m][r] \equiv Sx[m][r] - (Sx[m-1][r] - S'x[m-1][r]) w2[m]$
 である。

【0047】重み関数 $w1[n]$ 、 $w2[m]$ は、接続点での不連続を滑らかに補間する目的で用いる滑らかな関数であり、接続点で最大値をとり、音素区間の開始フレーム b 及び終了フレーム e で 0 となる。例えば、重み関数 $w1[n]$ 、 $w2[m]$ として、

$$w1[n] = \sin\{\pi(n-b)/2(c-b)\}$$

$$w2[m] = \cos\{\pi(m-c-d)/2(e-c-d)\}$$

を用いることができる。

【0048】以上が補間操作であり、これにより滑らかに補間された目的時間長スペクトルボタンが生成される。ちなみに、上述した基準スペクトルボタンの非線形伸縮及び補間操作は、処理 S120 にて生成された VCV 単位接続スペクトルボタンに対して行われる。これは先に個々の VCV 単位を伸縮すると VCV 単位の接続区間が不定となり、接続を行いにくくなるからである。

【0049】さて、1つの VCV 単位のスペクトルボタン内には、各母音、子音に対応した定常部分と、母音から子音、また子音から母音へといった遷移部分が含まれる。自然音声の観察から、発話速度を例えば速くしたとき、継続時間長の短縮に寄与するのは、主として定常部分であり、遷移部分は定常部分ほどには短縮されないという知見が得られる。これは例えば、スペクトログラムにおけるフォルマントの観察から理解される。スペクトログラムにおいては、定常部分はフォルマントがほぼ同一周波数位置に継続して現れるために時間軸に沿った一定のボタンとして観察され、遷移部分はフォルマントの周波数位置がある音素における位置から次の音素における位置へ移動するために変動ボタンとして観察される。上述した発話速度を速くした場合の現象は、定常部分が短縮され、遷移部分はそれほど変化しないといった形で観察される。スペクトルボタン生成部 56 により生成された目的時間長スペクトルボタンを用い、本装置から出力される合成音声は、このような自然音声の特徴を再現することができ、自然な合成音声を得ることができた。ちなみに、従来の合成音声装置は、VCV 単位内を一律に伸縮するため、例えば発話速度を速くした場合、遷移部分の短縮によりそのフォルマントの変化の傾きが保たれず大きくなり、自然音声の特徴の再現の程度が十分でなかった。

【0050】VCV 音節の継続時間長の変化は文全体の発話速度の変化以外に、音韻環境や隣接音韻の影響によっても変動する。上述したスペクトルボタンの伸縮の特徴は、このような要因による場合にも同様である。よって本装置は、音韻継続時間長生成部 54 にて、これら要因の影響を補償した音韻継続時間長を生成すれば、スベ

クトルボタン生成部 56 における規則上では特に意識することなく、これら要因の補償効果が反映されたスペクトルボタンを生成することができる。

【0051】なお、上述の例では、発話速度は 3 種類としたが、上述の説明から明らかなように本発明はその種類の数に限定されず、2 つ以上の異なる発話速度の代表時間長スペクトルボタンを用いて実施することができる。また、上述の例では、目的時間長スペクトルボタンを求めるために、その継続時間長を挟む大小の継続時間

10 長に対応した 2 つの代表時間長スペクトルボタンを選択し、それらの対応関係を用いた。この場合、上述したように目的時間長スペクトルボタンの各点の位置はそれに対応する 2 つの代表時間長スペクトルボタンの点の位置を継続時間長に応じて内分することにより定められる。しかし、3 つ以上の代表時間長スペクトルボタンを同時に用いて、目的時間長スペクトルボタンの精度の向上を図ることも理論上、可能である。例えば、3 つの代表時間長スペクトルボタン（継続時間長 D_i ($i = 1 \sim 3$)) の互いに対応する点を $P_{xyi}[p]$ とするとき、例えば 3 つの点 (D_i , $P_{xyi}[p]$) を通る 2 次関数を求め、その関数の目的継続時間長における値により、目的時間長スペクトルボタンに対応する DP バスを定義することにより継続時間長変化の 2 次項を考慮することができる。

【0052】スペクトルボタン生成部 56 は、上述したように各 VCV 音節毎に DP マッチングを行う構成としたが、例えば、VCV 音韻の種類毎に予め DP マッチングを行ってその DP バスをデータベースに格納しておき、音声合成時にはそれを読み出して用いるという構成も可能である。その構成によれば合成時における処理 S130 を省略することができ、装置の処理負荷を軽減することができる。また、さらに、DP バスのマクロバス、ミクロバスの分解も予め行うことも可能であり、ミクロバスをデータベースに格納して、音声合成時にこれを取り出して使用する構成とすることができる。

【0053】本装置は、VCV 音節という限定された単位のみに基づいてデータベースが構成されるので、従来の単位を拡大する方法に比べてデータベースの規模が小さくすることができる。しかも自然音声の有する異なる継続時間長間での非線形伸縮を反映した目的時間長スペクトルボタンが得られるので、自然な音韻の合成音声を実現される。

【0054】〔実施の形態 2〕本発明の第 2 の実施形態に係る日本語音声合成装置のブロック構成は、上記第 1 の実施形態の装置と同様であり、スペクトルボタン生成部以外の他の構成要素は同様の機能を有するので、説明を省略する。本装置のスペクトルボタン生成部に関しては、上記装置と異なるスペクトルボタン生成方法が採用されるため、区別のためその符号を変え、スペクトルボタン生成部 200 と記す。以下、スペクトルボタン生成

部200の処理を詳しく説明する。

【0055】スペクトルパタン生成部200はVCVデータベース202を有し、これにVCV単位のスベクトルパタン（代表時間長スペクトルパタン）が格納・保持されている。

【0056】VCVデータベース202は、VCV単位の各種類について基本的に1種類の代表的な発話速度についてのスペクトルパタン（代表時間長スペクトルパタン）しか保持しない。その代わりに、VCVデータベース202に保持された代表時間長スペクトルパタンに対する他の代表時間長スペクトルパタンのマイクロバス情報が、VCVデータベース202に予め格納・保持される。以上の点が上記実施形態のスペクトルパタン生成部56と異なる主な点である。

【0057】VCVデータベース202に格納する代表時間長スペクトルパタンの発話速度は、例えば6mora/秒といった低速であることが望ましい。その理由は、発話速度が速くなるにつれ、スペクトルパタンに含まれる音韻情報が欠落するからである。つまり情報が欠落した高速発声におけるスペクトルパタンから精度のよい低速発声におけるスペクトルパタンを生成することは難しいが、逆に低速発声におけるスペクトルパタンを捨象して精度のよい高速発声におけるスペクトルパタンを生成することは容易であるからである。

【0058】マイクロバスをVCVデータベース202に格納する方法について説明する。図5は、マイクロバスをVCVデータベース202に格納する方法を示す模式図である。以下、記号に関して上記実施形態と同一の表記を用いる。VCVデータベース202の作成時には、例えば3種類の発話速度（低速、中速、高速）で発声された音声のスペクトルパタン $Ss[i]$ 、 $Sm[j]$ 、 $Sf[k]$ が用意される。上述したように、これらのうち、最も発話速度の遅いスペクトルパタン $Ss[i]$ が、VCVデータベース202に格納される基準スペクトルパタンに選ばれる。この基準スペクトルパタン $Ss[i]$ に対する代表時間長スペクトルパタン $Sm[j]$ 、 $Sf[k]$ それぞれのDPバスが求められ（S210）、これを上記実施形態にて説明したように分解してマイクロバス $dPsm$ 、 $dPsf$ が生成される。そして、基準スペクトルパタン $Ss[i]$ の各フレームに対応するマイクロバスの値 $dPsm[i]$ 、 $dPsf[i]$ が求められる（S215）。これら各フレームに対応した複数のマイクロバスの値をマイクロバスベクトルと呼ぶことにする。なお、 $dPss$ は敢えて計算する必要はなく、全ての *

$$dPst2[i] = \{(Ds - Dt) / (Ds - Dm)\} dPsm2[i] \quad \dots\dots\dots (7)$$

であり、目的時間長 Dt が $Dm < Dt \leq Df$ を満たすときには、

$$dPst2[i] = dPsf2[i] + \{(Dt - Df) / (Dm - Df)\} (dPsm2[i] - dPsf2[i]) \quad \dots\dots\dots (8)$$

である。なお、ここで媒介変数として p の代わりにフレーム番号 i を用いた。

【0063】マクロバス $P'st[i]$ は、(4)式に表され

*フレームにおいて $dPss[i] = 0$ である。よって $dPss[i]$ をマイクロバスベクトルの成分に含めるかどうかは任意である。一般には、処理の都合上問題がなければ、データ量を抑制するため、 $dPss[i]$ はマイクロバスベクトルの成分には含めない。

【0059】これらマイクロバスベクトル（ $dPsm[i]$ 、 $dPsf[i]$ ）が基準スペクトルパタンの各フレームの情報と組み合わせられ（S220）、基準スペクトルパタンの各フレーム毎にVCVデータベース202に格納される。なお、スペクトルパタン生成部200は、スペクトルパタン生成処理において各代表時間長スペクトルパタンが発声された継続時間長 Ds 、 Dm 、 Df を必要とするため、これらを何らかの手段で保持する。これら継続時間長を基準スペクトルパタンの情報と組にして保持する構成とすることも可能である。

【0060】以上の手順により、基準スペクトルパタン $Ss[i]$ の各フレームに、発話速度に応じたマイクロバス変化の各サンプル点の情報が付加され、その基準スペクトルパタンがVCVデータベース202に音声合成処理に先だって格納される。スペクトルパタン生成部200は、このVCVデータベース202を用いて、音韻継続時間長生成部54から出力されるVCV音節の継続時間長に応じた目的時間長スペクトルパタン $St[h]$ の生成を行う。図6、図7は、スペクトルパタン生成部200の処理を説明する模式図である。

【0061】スペクトルパタン生成部200には、音韻継続時間長生成部54から音韻継続時間長を付与された音韻記号列が入力される。音韻記号列はVCV単位に分割される（S250）。当該VCV音節に対応する基準スペクトルパタンに関するデータが、VCVデータベース202から検索され取り出される（S255）。そして、各VCV音節の基準スペクトルパタン同士の接続が行われる（S260）。

【0062】次に、VCVデータベース202に格納されたマイクロバスベクトル300に基づいて、音韻継続時間長生成部54から与えられるVCV音節の継続時間長に応じたマイクロバス310（ $dPst$ ）の生成が行われる（S265）。このマイクロバスの合成は、例えば、各フレームのマイクロバスベクトルの成分を目的時間長に応じて線形補間して、各フレームに対応する目的時間長スペクトルパタンのマイクロバスの値を定めることにより行われる。具体的には、目的時間長 Dt が $Ds \leq Dt \leq Dm$ を満たすときには、

るように継続時間長の比を傾きとする直線である。よって、マクロバス315として傾き（ Dt/Ds ）を有する直線が生成される。このマクロバス $P'st$ にマイクロバス d

17

Pstを加算することにより、目的時間長スペクトルボタンに対するDPバス320 (Pst) が生成される (S270)。また、ここで上記実施形態同様、DPバスのフレームの格子への吸着が行われ、それに対応して、やはり上記実施形態同様、フレームの間引き・伸長処理が行われるとともに、補間処理によるスムージング処理が行われる (S275)。

【0064】以上のスペクトルボタン生成部200における処理により得られたスペクトルボタンを用いて、音声合成部64にて音声の合成が行われる。

【0065】第1の実施形態の装置は保持するスペクトルボタンがVCV単位のみに限定される分、データベースを小さくできた。しかも、目的時間長に最も近い時間長のスペクトルボタンを伸縮することにより精度のよい目的時間長スペクトルボタンが得られる。

【0066】第2の実施形態に係る本装置は、第一の装置が3つの発話速度に対応したスペクトルボタンをデータベースに保持していたのに対し、保持するスペクトルボタンを単一の発話速度のものに限定することにより一層のデータベースの縮小を図ることができる。具体的には、本装置において、VCVデータベース202に保持する必要があるのは基本的には1つの発話速度に対応したスペクトルボタンとマイクロバスベクトルである。例えば、スペクトルボタンの1フレームのデータは19個のパラメータで構成され、マイクロバスベクトルを構成するパラメータは上述の例では最低2つとすることができる。よって、マイクロバスベクトルによるデータ量の増加は、スペクトルボタンを3つから1つにすることによるデータ量の削減に比べればわずかなものであり、VCVデータベース202に格納されるデータ量を大幅に抑制することができる。

【0067】また、VCV単位の前データが代表時間長の種類に比例して増加すると、それらの接続における組合せ数も増加する。そのため、それぞれの接続の滑らかさを確保するためには、データ収録を慎重に行わなければならない、データベース作成の労力が増大する。このVCV単位の接続に関しても、本装置では1つの発話速度のVCV単位間でのみ接続が行われるため、その組合せも大幅に減少し、データ収録、データベース作成の

18

労力を低減することができる。よって本装置によれば、非線形伸縮による自然な音声の合成を可能としつつ、かつ簡便に合成音声装置を構成することができる。

【0068】

【発明の効果】本発明の音声合成装置によれば、データベースを小規模に抑制しつつ、かつ任意の目的発話速度に対して、接続歪みや音色の不自然な変化が抑制されたスペクトルボタンが生成され、自然な音韻の合成音声を生成することができるという効果が得られる。

10 【図面の簡単な説明】

【図1】 本発明に係る日本語音声合成装置のブロック構成図である。

【図2】 第1の実施形態に係る音声合成装置のスペクトルボタン生成部の処理を説明する模式図である。

【図3】 同一のVCV音節に対応した3種類の発話速度の代表時間長スペクトルボタンをそれぞれ示す模式図である。

【図4】 DPバスを示す模式図である。

20 【図5】 第2の実施形態に係る音声合成装置において、マイクロバスをVCVデータベースに格納する方法を示す模式図である。

【図6】 第2の実施形態に係る音声合成装置のスペクトルボタン生成部の処理を説明する模式図である。

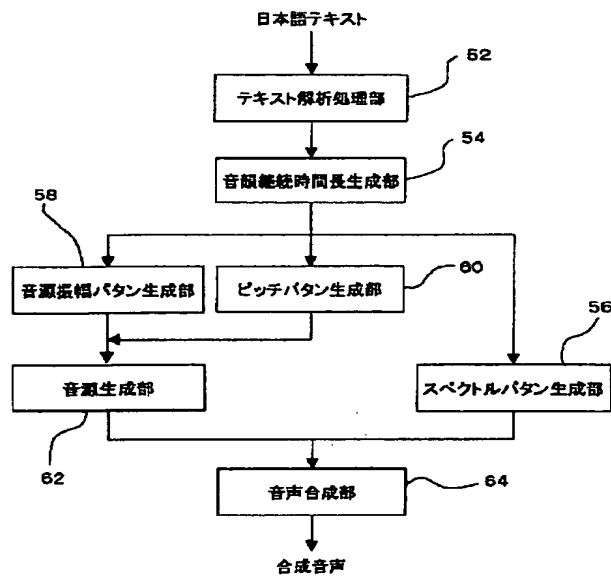
【図7】 第2の実施形態に係る音声合成装置のスペクトルボタン生成部によるDPバスの合成処理を説明する模式図である。

【図8】 従来のスペクトルボタン生成方式における処理の流れを示す模式図である。

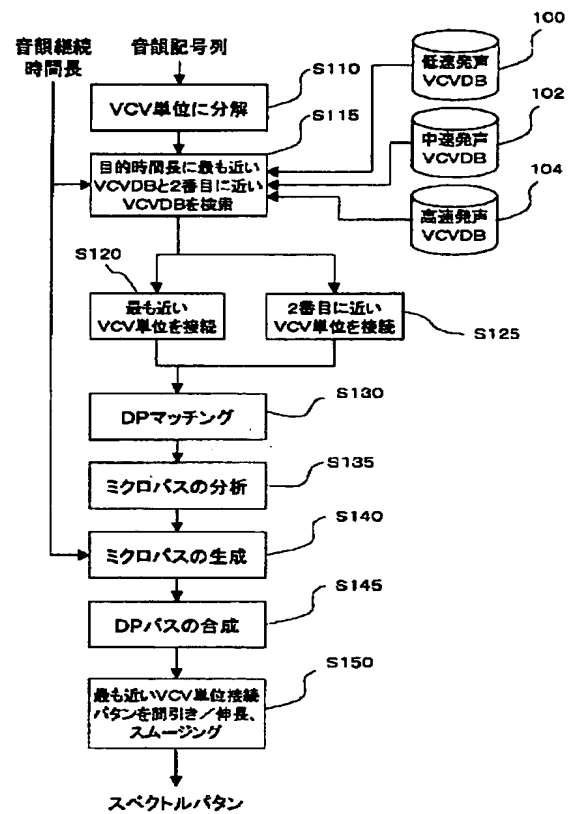
【符号の説明】

30 52 テキスト解析処理部、54 音韻継続時間長生成部、56、200 スペクトルボタン生成部、58 音源振幅ボタン生成部、60 ピッチボタン生成部、62 音源生成部、64 音声合成部、100 低速発声VCVデータベース、102 中速発声VCVデータベース、104 高速発声VCVデータベース、202 VCVデータベース、300 マイクロバスベクトル、310 マイクロバス、315 マクロバス、320 DPバス。

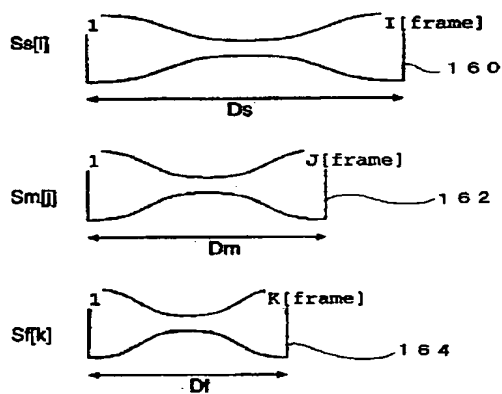
【図1】



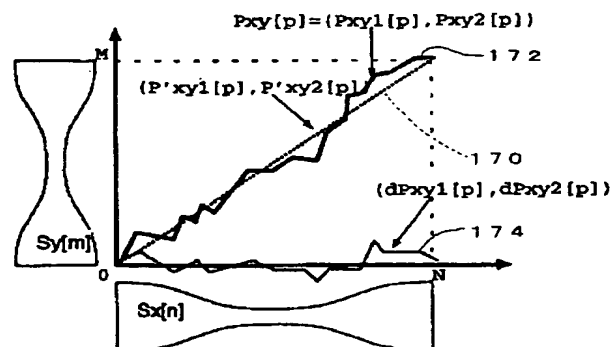
【図2】



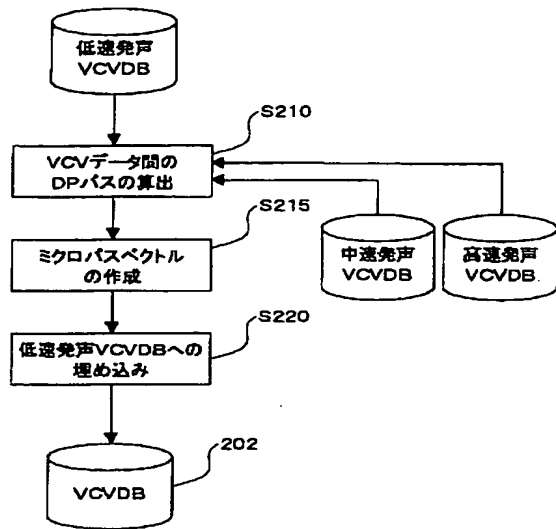
【図3】



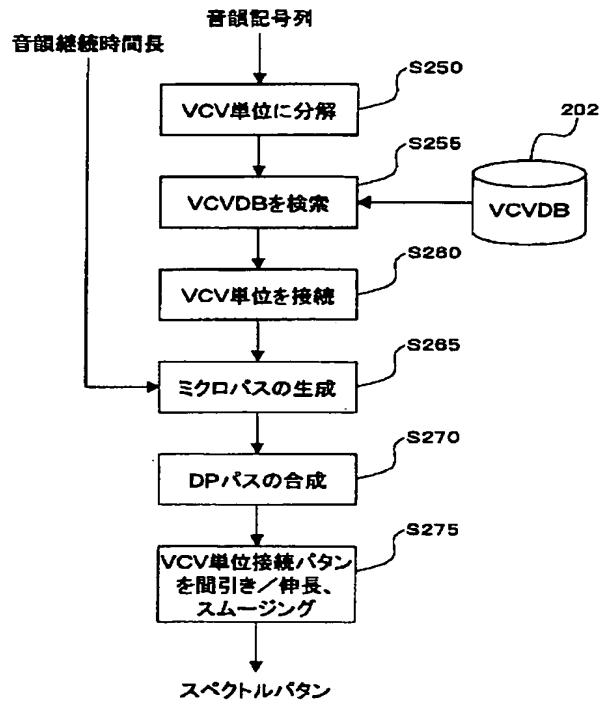
【図4】



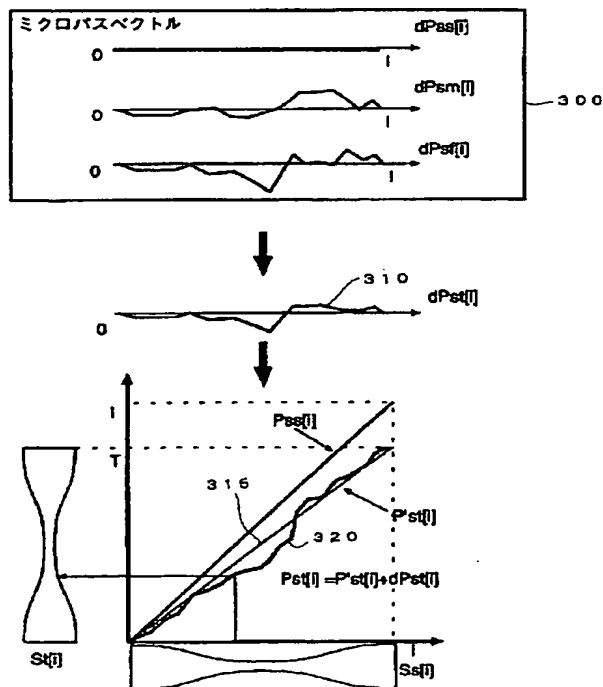
【図5】



【図6】



【図7】



【図8】

